Reliability-Metric Varieties and Their Relationships

Alan P. Wood • Tandem Business Unit • Compaq Computer Corporation • Cupertino

Key Words: Commercial reliability program, Computer system, Field failure data, Reliability - field, Reliability metrics, Part replacement.

SUMMARY & CONCLUSIONS

Although reliability theory is based on the concept of failures, the definition of a failure may be very different from different perspectives. To a hardware engineer, a failure means a part replacement and verification of the failure. To a manufacturing repair depot, a failure is a returned part. To the finance department, a failure is a warranty claim. To a service organization, a failure is a service call for corrective maintenance (as opposed to planned or preventive maintenance). To a customer, a failure is a degradation of service or capability of the product. These various definitions of failure lead to various types of reliability metrics, such as part replacement rate or service call rate, in addition to the classic reliability metrics such as constant failure rate.

This paper describes many types of reliability metrics and their interrelationships, both in theory and in practice. For the practical metric relationships, we draw from our experience with many reliability metrics at the Tandem Business Unit of the Compaq Computer Corporation. Our main conclusions are

- The classical engineering view of failures is not sufficient for customer satisfaction with reliability in our market.
- Multiple reliability metrics need to be tracked to satisfy the needs of various organizations.
- Many different functional groups must work together to ensure customer satisfaction with reliability.

1. INTRODUCTION

The Tandem Business Unit of the Compaq Computer Corporation designs and manufactures fault-tolerant computers for business-critical applications. These NonStop™ systems are the dominant computers used to run stock exchanges and automated teller machines. Our systems include cabinets with power and cooling subsystems, processor boards, communications boards, storage devices such disk and tape drives, and peripheral devices such as printers and network interface devices. Our customers are the end-users of the computers and demand the highest levels of reliability and availability, where those demands are based on the customers' perspective of failure.

Compaq performs almost all the service for our NonStop systems. We use a very sophisticated, worldwide problem reporting and tracking system that allows us to track every customer call pertaining to our hardware and software. Different codes allow us to distinguish among questions, administrative requests, and problems, and to differentiate among hardware, software, and other types of problems. Hardware replacements and non-replacement hardware corrective actions are tracked. By combining this data with our installation data base and manufacturing repair data, we are able to create a set of metrics that provides different views of reliability. A more complete description of our reliability program activities and data tracking capabilities are provided in Ref. 1.

This paper presents the reliability metric suite we use. Section 2 describes many different reliability metrics, segregated into constant rate metrics and probability of success metrics. Section 3 presents the set of metrics we use and the relationship among those metrics, including the required involvement of various organizations in order to improve the customer perspective of reliability. Section 4 presents the various types of reporting we perform for the metrics. Section 5 describes our experiences using the metrics, showing that the expected theoretical relationships do not always exist in practice.

2. TYPES OF RELIABILITY METRICS

Reliability metrics are grouped into two general categories of metrics in this paper: constant rate metrics (exponential distribution) and probability of success metrics (non-exponential distribution). The most commonly used reliability metrics are constant rate metrics such as a constant failure rate. These metrics are generally considered a good approximation for the middle section of the bathtub curve. Mean life metrics such as MTBF usually assume (explicitly or implicitly) an exponential distribution, which makes them equivalent to constant rate metrics. Non-exponential reliability distributions such as the Weibull distribution require at least 2 parameters, so a single value such as MTBF is not sufficient to characterize the distribution. For those distributions, a more appropriate metric is a measure of mission success such as reliability. This metric will usually be time dependent, e.g., the probability of mission success will vary depending on the length of the mission.

0-7803-6615-8/01/$10.00 © 2001 IEEE

Constant Rate Reliability Metrics

Constant rate reliability metrics are widely used throughout the electronics industry. These metrics may be selected because:

1) they are good approximations of the reliability behavior of a product during its "useful life", or
2) they are the only metrics that can be measured from field data, or
3) they are simple to calculate and easy to explain, which is important for presentations to executives.

If the constant rate is represented by the parameter λ, the mean value of the exponential distribution is $1/\lambda$. Therefore, constant rate metrics can either be described as a rate or as a mean life, e.g., constant failure rate or MTBF. Table 1 contains a long, but certainly not exhaustive, list of constant rate metrics and their equivalent mean life inverses along with an indication of when the metrics might be appropriate.

Constant rate reliability metrics are sometimes used as surrogates for time-dependent metrics by specifying different failure rates for different periods of time such as the Infant Mortality phase. For example, a failure rate goal might be stated as 3λ for first 3 months of production or operation, 2λ for next 3 months, and λ after 6 months. Another approach for replacing a time-dependent metric with a constant failure rate is to determine the expected number of failures for a certain time period and to specify a constant failure rate during that time period. For example, a product that follows the bathtub curve might be expect to have 50 failures in a population of 1,000 in the first year, 30 failures in each of the next 3 years, and 60 failures in the 5th year. This product's reliability may be approximated with a constant failure rate of 4.6 failures per million hours (4,600 FITs) for 5 years, equivalent to 200 failures in a population of 1,000 in 5 years.

Constant Rate Metric	Mean Life Equivalent	Definition	Use
Failure rate	Mean-Time-Between/Before-Failure (MTBF) or Mean-Time-To-Failure (MTTF)	Total failures divided by total population operating time; can be expressed as failures per hour, failures per year (AFR=annualized failure rate), failures per million hours, failures per billion hours (FITs), etc.	Standard metric for reliability predictions; measure of inherent product hardware reliability
Failure rate using cycles instead of time	Mean-Cycles-Between/Before-Failure (MCBF) or MCTF	Total failures divided by total population number of cycles	Standard metric for reliability predictions when usage is more relevant than time
Failure rate using distance instead of time	Mean-Miles-Between/Before-Failure (MMBF) or MMTF	Total failures divided by total population number of miles	Standard metric for reliability predictions when usage in terms of distance is more relevant than time, e.g., automobiles or aircraft
Part Return/Repair Rate	MTBPR (R= return/repair)	Total parts returned/repaired divided by total population operating time	Useful for sizing a repair depot or manufacturing repair line
Part Replacement Rate	MTBPR (R=replacement)	Total parts replaced divided by total population operating time	Used as surrogate for failure rate when no failure analysis is available; useful for warranty analysis
Service or Customer Call Rate	MTBSC	Total service/customer calls divided by total population operating time	May be customer perception of failure rate; useful for sizing support requirements
Warranty Claim Rate	MTBWC	Total warranty claims divided by warranted population operating time	Useful for pricing warranties and setting warranty reserve
Service Interruption Rate	MTBSI	Total service interruptions divided by total population operating time	May be customer perception of failure rate; may be an availability metric
Maintenance Action Rate	MTBMA	Total maintenance actions divided by total population operating time; may include preventive maintenance	May be customer perception of failure rate; useful for sizing support requirements

Table 1. Constant Rate Reliability Metrics

Probability of Success Metrics

When reliability is not exponentially distributed, i.e., when the failure rate is not constant, constant rate metrics do not typically apply. Also, although non-exponential distributions have a mean so that mean-time-to-failure can be calculated, it may be confusing to talk about an MTBF or MTTF because those metrics are normally associated with a constant failure rate. A more useful metric for non-exponential distributions is reliability, defined as the probability of mission success or, more formally, as the probability that an item performs a required function under stated conditions for a stated period of time.

A different way to specify a probability of success is as a percentage of a population that survives a specified duration. For this metric, percentiles of the distribution may be used. These percentiles are the amount of the product population expected to fail by a certain time and are stated as a "B" life (commonly used in the automotive industry) or an "L" life (commonly used for mechanical devices such as fans). For example, an L10 life of 300,000 hours means that 10% of the product population will have experienced a failure after 300,000 hours of operation, and a B50 life of 5 years means that 50% of the product population will have experienced a failure after 5 years of operation. Some metrics may be stated with a confidence level, e.g., R96C90 in the automotive industry means a reliability of 96% with 90% confidence.

Probability of success metrics can use the same variations on the definition of failure as constant rate metrics. For example, failure can be replaced with part replacement, service interruption, or maintenance action when deriving a reliability distribution.

Measuring a probability of success metric from field reliability data requires more information than measuring a constant rate metric. For a constant rate metric, it is only necessary to know the field population operating time and number of failures (or number of removals, service calls, etc.). The population operating time is often estimated from shipment records and knowledge of standard installation and operating procedures. A probability of success metric requires that installation, operating profiles, and failure times be available for each individual product in the field population. This requires separate field tracking, usually by serial number, for each individual unit.

3. RELIABILITY METRIC RELATIONSHIPS

For our NonStop systems, Compaq measures several different reliability metrics to reflect the different views of reliability. Figure 1 shows our suite of reliability metrics and the theoretical relationship among them. Note that the classic engineering view of reliability (failure rate) may be significantly different than a customer's view of reliability (hardware problem call rate or corrective maintenance rate) and other organizations' view of reliability. Our executives are most interested in the corrective maintenance rate and hardware problem call rate since those are the customer view of hardware reliability. Our manufacturing organization is most interested in the part return rate for sizing the repair lines. Our logistics organization is most interested in the part replacement rate for spare parts provisioning.

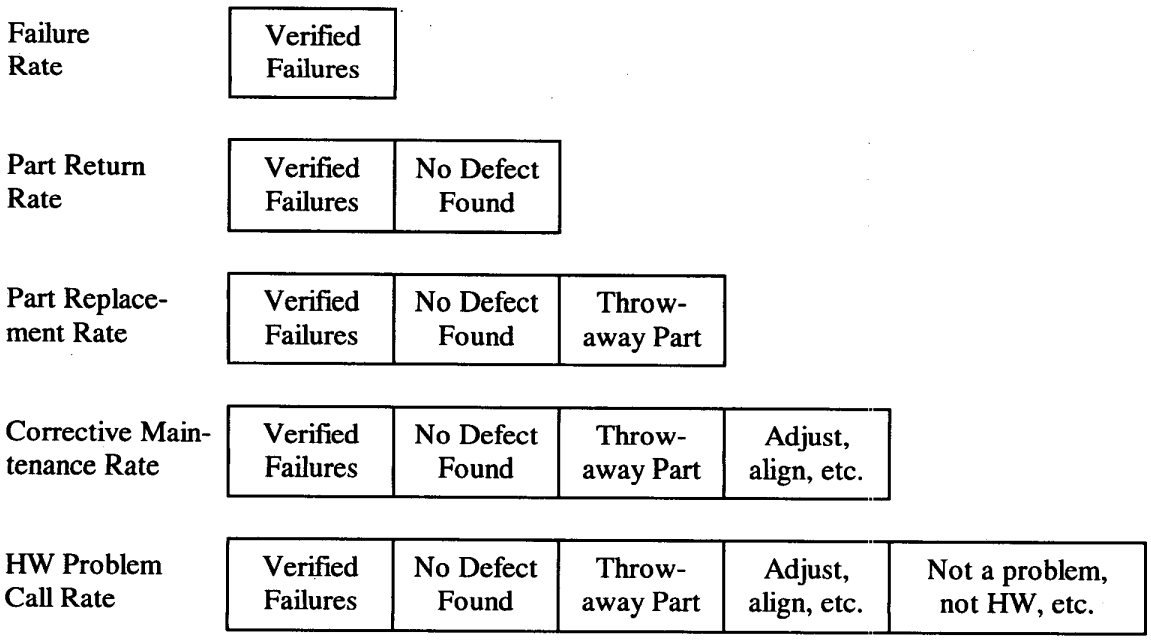

Figure 1. Example Reliability Metric Relationship

Most of our products are high-value circuit boards, so the part replacement rate and part return rate are identical. Examples of products for which the replacement rate and return rate may be significantly different are tape drives and printers. These peripheral devices often have small mechanical devices that wear out or become misaligned and need to be replaced. The small mechanical devices are not returned, but if the entire printer or tape drive needed to be replaced for repair beyond the capability of service personnel, the printer or tape drive would be returned.

Tracking these different sets of metrics is very useful for understanding the basis for customer complaints. If the executives say we have customers complaining about a certain product, but the engineers say the product failure rate is no different than other similar products, we can examine the other metrics. There may be diagnostic or support issues that are causing excessive removals. There may be mechanical parts that get out of alignment easily but do not need to be replaced. The service log may have confusing messages that cause the customer to call when there is no problem. All of these problems need to be fixed to improve customer satisfaction, but failure rate does not provide the appropriate visibility into the problems.

As shown in Figure 1, the customer perception of reliability may have many components. Even if engineering reduces the number of verified failures to a negligible number, the customer may not be happy with the product reliability. Different functional groups may need to be involved to fix issues that the customers perceive as reliability problems. Our suite of metrics allows us to determine the appropriate issues and the appropriate functional groups that need to be involved in reliability improvement activity.

If the return rate is significantly higher than the failure rate, meaning the "No Defect Found" rate is high, engineering and manufacturing need to work with the service organization to determine why parts that appear to be good are being returned. This may result in changes to diagnostics, manuals, service training, manufacturing testing, or even to the failure symptom data recorded by the product. If the replacement rate is high, but the parts are throwaway items, engineering and manufacturing may work with the service organization to modify normal service practices and obtain some parts for analysis. If the corrective maintenance rate is significantly higher than the replacement rate, engineering needs to work with the support organization to determine the service actions being performed. There may be product or maintenance changes that could prevent frequent adjustment, alignment, and other non-replacement actions. If the hardware problem call rate is significantly higher than the corrective maintenance rate, hardware problems are probably being confused with software problems, non-problems, or administrative issues. Engineering needs to work with the support organization to determine the cause of the confusion.

Our suite of metrics also allows us to track the impact of product design or maintenance policy choices on customer-perceived reliability. Since our customers demand high availability, we may choose to quickly replace parts rather than to spend additional time troubleshooting. This may cause higher repeat service calls or higher replacement rates in order to reduce repair time. We may choose to bundle multiple hardware items into a single replaceable unit in an attempt to ensure that the failed part is correctly identified and removed more often. This should show up as reduced service calls that offsets the additional expense involved in stocking and transporting the more expensive bundled items.

Although this paper is limited to hardware reliability, the same concepts apply to system reliability, software reliability, or even logistics metrics such as dead-on-arrival (DOA). For example, the customer perception of a software failure may include operational errors using the software, in addition to the actual software coding and documentation errors. The appropriate software reliability metric might be calls to a help desk rather than unique software defects. Reducing calls to a help desk may require creation of special wizards, toolbars, and so forth in addition to fixing actual software defects.

4. RELIABILITY METRIC REPORTING

To ensure that our non-failure rate metrics accurately reflect product reliability, we find it necessary to carefully audit the service call data. For example, if a processor stops and a customer calls in with a problem, it would normally be designated a hardware problem. However, after a support analyst evaluates the problem, they may find that the problem is a software problem. Since the original problem designation is hardware, we would end up with incorrect metrics if we did not audit the data. Also, in some cases, parts may be removed for software problems, an intermittent hardware problem may appear to be a software problem, customer questions may turn out to be hardware problems, incorrect operation may be disguised as hardware problems, and so forth. We read the text of each service call and reclassify them when appropriate.

We use various types of charts and graphs to report our set of metrics. An example is shown in Figure 2. As changes are made to a product, it may be important to track different product revisions to highlight the different field reliability of those revisions. Figure 2 shows a product that exhibited several failure mechanisms that had not been anticipated during the pre-production stages. Corrective action included re-specification of a critical component, layout improvement, and firmware updates. The new revision was tracked separately and the difference was dramatically demonstrated in the field performance data. The old version continued to exhibit its unsatisfactory failure rate, but the new version was immediately more reliable and quickly settled down to its steady state value, where it remains today. Note that the metric in Figure 2 is the replacement rate, even though the product changes impacted only the failure rate. It was possible to demonstrate the improvement more quickly from the replacement rate because it does not require failure rate verification. This is an example of the benefits of using multiple metrics.

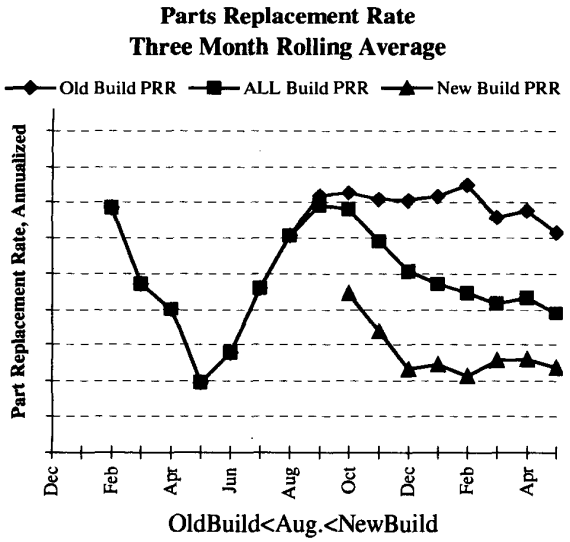

Figure 2. Example Constant Rate Metric Reporting

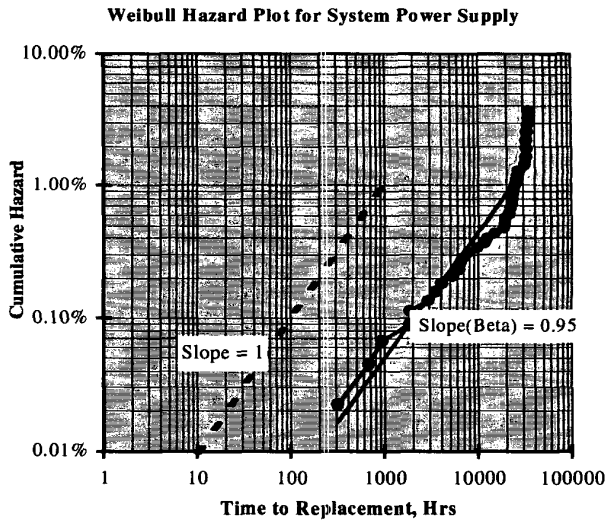

Figure 3. Example Weibull Hazard Plot

Similar types of reporting may be required to highlight the non-uniformity of a customer base. Different operating environments, applications, and operator experience may all contribute to differences in customer-perceived reliability. For examples, a more severe operating environment may increase the product failure rate. Less-experienced operators may have a higher problem call rate relative to more-experienced operators, even though the corrective maintenance rate is the same for both.

Some of our products such as power supplies, disk drives and fans may exhibit wear-out failure mechanisms. For these products, constant rate metric reporting is not sufficient to provide early detection of potential wear-out problems. Weibull hazard analysis is performed on a regular basis for these types of products. An example of a Weibull hazard plot for power supply replacements is shown in Figure 3. This plot is a log-log graph of cumulative hazard rate against run time of a particular power supply, showing significant premature wear-out. This power supply began to show signs of wear-out at about 20,000 hours. The trend quickly became very obvious, and root cause analysis was begun. Again it was possible to use replacements rates rather than failure rates to identify and track a product reliability problem.

5. RESULTS FROM RELIABILITY METRICS

We have found that customer-focused reliability metrics provide more timely visibility into field reliability problems than failure rate. Our field tracking system allows us to create metrics based on site visits or part replacements in advance of the time that the parts are returned for analysis. In addition we do not need to wait for the potentially lengthy failure analysis to complete in order to spot trends based on symptoms recorded during the site visits. Section 4 provided two examples of using a part replacement metric to identify a failure rate problem, which allowed us to take earlier corrective action.

There is an interesting relationship between part replacement rate and corrective maintenance rate. In theory, there are more corrective maintenance actions than part replacements because reseating a circuit board may repair transient errors and mechanical parts may require adjustment or alignment rather than replacement. For many of our parts, about 10% of corrective maintenance actions are non-part replacements. However, we have also found that multiple parts may be replaced on a single call either because of multiple failures (rare) or inadequate diagnostics or training. These extra part replacements almost exactly cancel the non-part replacement maintenance actions, so part replacement rate and corrective maintenance rate are essentially identical.

There is another interesting relationship between part return rate and failure rate. As our engineering group performs root cause analysis/corrective action and fixes problems found in the field, successive product revisions have decreased field failure rates. The part return rates also decrease, but the decrease is more than expected. The ratio between part return rates and failure rates seems to remain constant as failure rates decrease. This means that the number of parts removed and identified as "No Defect Found" are also decreasing, even though engineering has done nothing to specifically reduce the number of No Defect Founds. Our working hypothesis for this puzzling phenomenon is that many of the verified failures can also instantiate themselves as transient or intermittent failures, which makes them very difficult to find. Thus, fixing the verified failures is also fixing many of the No Defect Founds.

The same relationship seems to exist between corrective maintenance rate and failure rate as shown in Figure 4. These rates seem to decrease in concert when failure rate problems are fixed, even though we may have done nothing to decrease the other components of corrective maintenance rate. The transient or intermittent failures mentioned in the previous paragraph may cause a customer to reboot rather than request a part replacement. Thus, fixing the verified failures is also fixing many of the reboots.

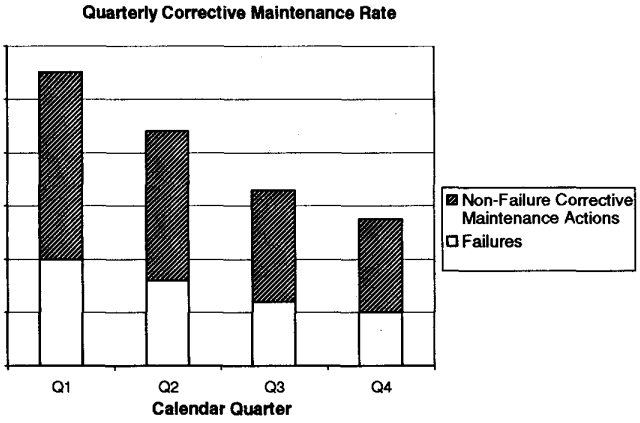

Figure 4. Comparing Corrective Maintenance Rates and Failure Rates

REFERENCES

1. J. G. Elerath, A. P. Wood, D. Christiansen, M. Hurst-Hopf, "Reliability Management and Engineering in a Commercial Computer Environment", *Proc. Ann. Reliability & Maintainability Symp.*, 1999 Jan, pp 323-329.

BIOGRAPHY

Alan P. Wood, PhD
Tandem Business Unit, Compaq
10300 N. Tantau Ave CAC05-52
Cupertino, CA 95014 USA

alan.wood@compaq.com

Alan Wood is the Director of the Reliability Engineering Department at the Tandem Business Unit of Compaq Computer Corporation. He has 20 years experience in reliability engineering for nuclear energy, defense, and commercial environments. He has a PhD in Operations Research from Stanford University and has published over 25 papers in the areas of reliability and availability. His current primary research interests are software reliability and computer systems availability. He is a member of IEEE and INFORMS and the IEEE 1413 Working Group.